EFFECT OF SHARED INFORMATION ON TRUST AND RELIANCE IN A DEMAND FORECASTING TASK

Ji Gao[1], John D. Lee[2]
Intel Corporation, Hillsboro, OR[1]
University of Iowa, Iowa City, IA[2]

People have difficulty relying on forecasting systems appropriately, which can lead to huge business losses. Sharing information regarding the performance of forecasting systems may lead to more appropriate trust and reliance. This study considered imperfect forecasting systems and investigated how sharing such information influences people's trust and reliance. A simulated demand forecasting task required participants to provide an initial forecast, select and view a model forecast, and then determine their final forecast. Results showed that participants' reliance on a forecasting model strongly depended on their trust in the model, which was often inappropriate. With shared information, participants' reliance was more sensitive to changes of their trust in the model. However, when the shared information exposed instances of poor performance of the model, it diminished compliance with the selected model forecast, which undermined the accuracy of the final forecasts. These results suggest that sharing information may promote more appropriate reliance in situations in which people over trust automation, but not in situations in which people tend to under trust automation.

INTRODUCTION

Demand forecasting is a task that strongly influences success in supply chains. Inappropriate reliance on forecasting systems can have negative impact on the company (Lee & Gao, In press). For example, Cisco's sales plunged 30 percent and wasted an inventory worth $2.2 billion because their decision makers failed to forecast a slow down in demand due to their over-reliance on their forecasting tool (Berinato, 2001; Paul, 2002). Similarly, the Nike over-relied on a forecasting system to order $90 million of shoes, which resulted in substantial excess inventory (Crane, 2001; Sterlicchi, 2003).

Many studies have shown that human's inappropriate reliance on automation can degrade performance (Lee & See, 2004; Mosier, Skitka, Heers, & Burdick, 1998; Parasuraman & Riley, 1997). Similarly with forecasting systems, people have considerable difficulty in appropriately reacting to the reliability of the forecasts provided by forecasting models (Goodwin, 2000; Goodwin & Fildes, 1999; Lawrence & Sim, 1999; Lim & Oconnor, 1996; Lim & O'Connor, 1995). People tend to rely heavily on their judgmental forecasts, even when the statistical forecasts are highly reliable (Lawrence, Goodwin, & Fildes, 2002; Lawrence, O'Connor, & Edmundson, 2000). Therefore, identifying ways to help forecasters rely on forecasting automation more appropriately is critical to improve forecasting performance.

Sharing information regarding the performance of forecasting models with other users may help users evaluate the capability of forecasting models and therefore lead to more appropriate trust and reliance. In this study, we considered two imperfect forecasting systems and investigated whether sharing such information improves appropriate trust and reliance. The ultimate goal is to identify conditions in which sharing such information promotes more appropriate trust and reliance.

METHOD

Participants

Twenty-four (17 male and 7 female, age Mean=22, age SD=2.6) engineering students at the University of Iowa participated in this experiment. It took each participant about two hours to finish the session and the average compensation was $20.53/participant.

Forecasting task

Participants interacted with a supply chain microworld in which they forecast demand for each of 75 trials. Figure 1 shows the screenshot of the microworld interface with labels. Their objective was to maximize forecast accuracy. For each trial, the participants viewed the history of demand, the history and current values of three factors influencing demand and then made an initial forecast. They then selected a forecasting model and adjusted the initial forecast to generate the final forecast.

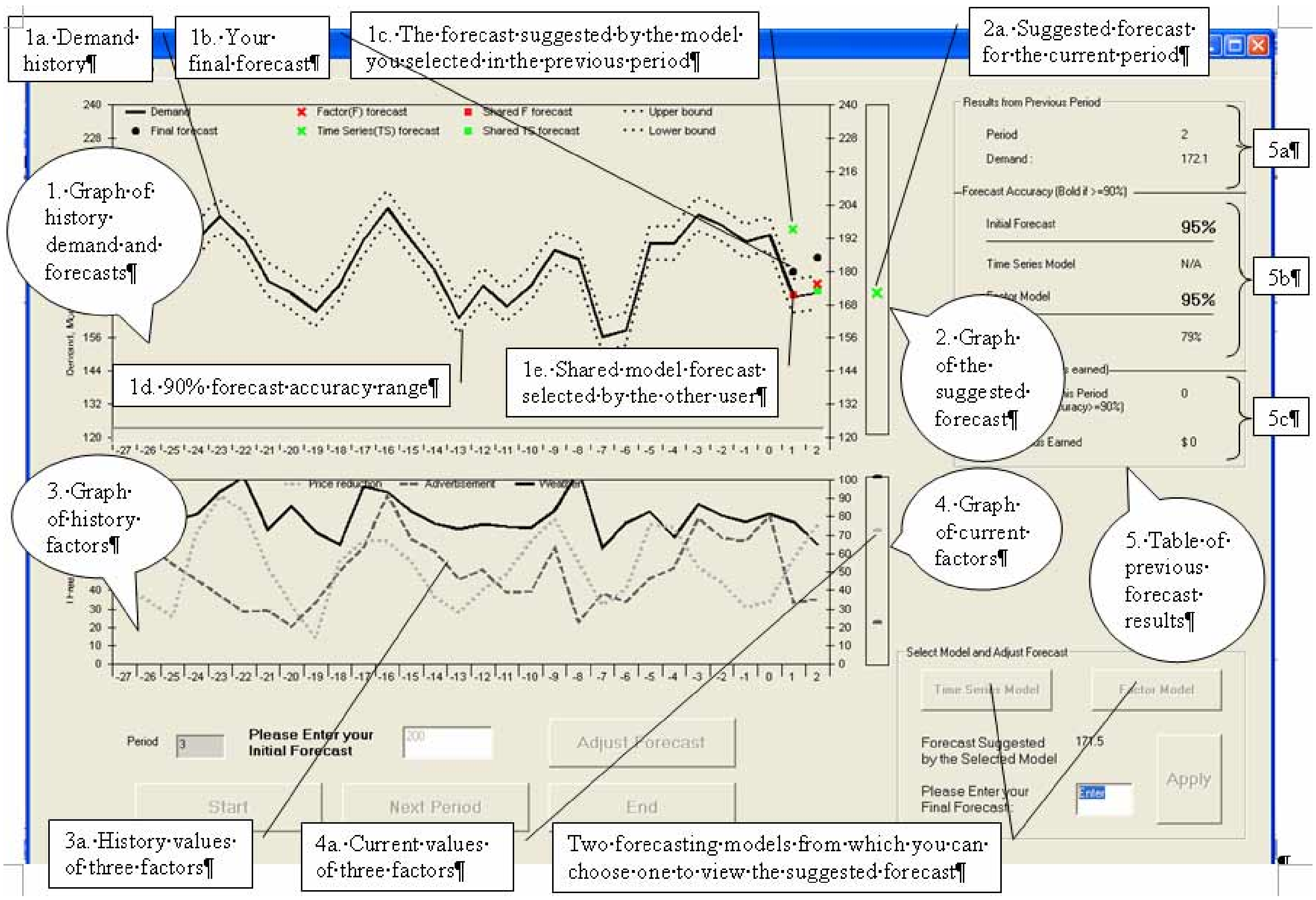

Figure 1. Screenshot and labels for the main interface of the demand forecasting microworld.

Table 1. Experimental design

		Block 1 (Trial 1-15)	Block 2 (Trial 16-30)	Block 3 (Trial 31-45)	Block 4 (Trial 46-60)	Block 5 (Trial 61-75)
No Sharing	Fault order 1 (6 subj.)	F>TS	TS>F	Both unreliable	F>TS	TS>F
	Fault order 2 (6 subj.)	TS>F	F>TS	Both unreliable	TS>F	F>TS
Share	Fault order 1 (6 subj.)	F>TS	TS>F	Both unreliable	F>TS	TS>F
	Fault order 2 (6 subj.)	TS>F	F>TS	Both unreliable	TS>F	F>TS

Two commonly used forecasting models were provided (Jarrett, 1990): Factor Model (FM) and Time Series Model (TSM). The FM predicted demand based on the variation of external factors and the TSM predicted demands based on the past history of demand. The participant could only view the forecast from one forecasting model for each trial. Participants in the information sharing condition were told that at the end of each trial, the model forecast selected by another user was shared in the graph that displayed the history demands and forecasts. The information regarding the use of the model by another user was pre-defined by computer program to simulate the other user. A sequence of alternating reliance on TSM and FM starting with reliance on TSM (i.e., TSM FM TSM FM….TSM FM) simulated the pattern of reliance of the other user.

Experimental design

The experiment used a mixed design: 2 (information sharing, between-subject) x 2 (fault order, between-subject) x 3 (demand profile, within-subject) x 5 (time blocks, within-subject), as shown in Table 1.

Three levels of demand profile corresponded to three levels of reliability of forecasting models: (1) dramatically changing demand, during which FM was more reliable than TSM (F>TS), (2) smoothly changing demand, during which TSM was more reliable than FM (TS>F), and (3) demand had elements of both smooth and dramatic changes when both TSM and FM were unreliable. Averaging over all the blocks, FM and TSM were equally reliable.

Procedure

Each participant completed an inter-personal trust questionnaire (Rotter, 1967), read written instructions concerning forecasting and forecasting models, and took a quiz to ensure they understood the task. The participants then completed in a practice session of 20 trials and the experimental session of 75 trials.

Each trial consisted of four steps: enter initial forecast, select one forecasting model (TSM or FM) to view the model forecast, adjust the initial forecast, and enter the final forecast. Before they observed the actual demand, they rated their trust in each forecasting model and their self-confidence in their initial forecast. Each participant was paid \$10 for completing 75 trials and could obtain a bonus of \$0.25 for each trial if the final forecast accuracy for the trial was equal or greater than 90%.

Dependent variables

Three categories of dependent variables were used in this experiment: (1) forecasting task performance variables including *initial and final forecast accuracy*, (2) decision process variables including *reliance on model* and *compliance with the model forecast*, and (3) subjective measures including the participant's *trust in each model* and *self-confidence in his/her initial forecast*.

The forecast accuracy was calculated by

$$Accuracy = \left(1 - \frac{Absolute\ Error}{60}\right) \times 100\% \quad (1)$$

Reliance on a model was coded as 1 if the model was selected and 0 if not selected. Appropriate reliance was a binary variable with 1 indicating that the model with the higher reliability was selected and 0 representing that the model with the lower reliability was selected. The compliance with the model forecast was calculated by

$$Compliance = \frac{|FF - IF| - |FF - MF|}{|MF - IF|} \quad (2)$$

where FF, IF, and MF denote the final, initial, and model forecast, respectively. Equation (2) was developed so that compliance represents how relatively close the final forecast is to model forecast compared to the participants' initial forecast. Compliance greater (less) than zero represents a situation where participants depended more (less) on the model forecasts and less (more) on their initial forecasts to generate their final forecasts.

Participants rated their trust and self-confidence on a scale of 0 to 10, with 10 being the highest.

Data analysis

The data were aggregated for each block and the block-based analysis was performed for blocks 1, 2, 4, and 5. Repetition was defined with two levels: the first half (blocks 1 and 2) and the second half of the experiment (blocks 4 and 5). Specifically, a 2 (information sharing) x 2 (fault order) x 2 (demand profile) x 2 (repetition) repeated-measures analysis of variance (ANOVA) was performed. There was no difference in participants' interpersonal trust scores for the not-sharing (Mean=59.83, SD=12.19) and sharing (Mean=57.41, SD=11.92) groups, p =0.94.

RESULTS

Participants' initial forecasts were significantly more accurate when demand was smoothly changing than when demand was dramatically changing. The mean accuracy of participants' initial forecasts was 71% when demand was dramatically changing and was 82% when demand was smoothly changing, F (1, 20) = 54.08, $p <$ 0.0001.

Although FM and TSM were equally reliable overall, participants showed a bias towards selecting FM. Overall, participants selected the FM 56% of the time and selected the TSM 44% of the time. Participants selected the FM 71% of the time when FM was more reliable, whereas they selected the TSM only 60% of the time when TSM was more reliable. That is, the mean appropriate reliance was 71% when demand was dramatically changing and 60% when demand was smoothly changing, F (1, 20) = 5.1, p = 0.0353, leading to an overall appropriate reliance of 65.5%. Sharing information had no statistically significant effect on the degree of appropriate reliance.

Information sharing did affect compliance with the model forecast. Figure 2 shows the three-way interaction between sharing information, demand profile, and repetition, F (1, 20) = 5.00, p = 0.0369. Sharing

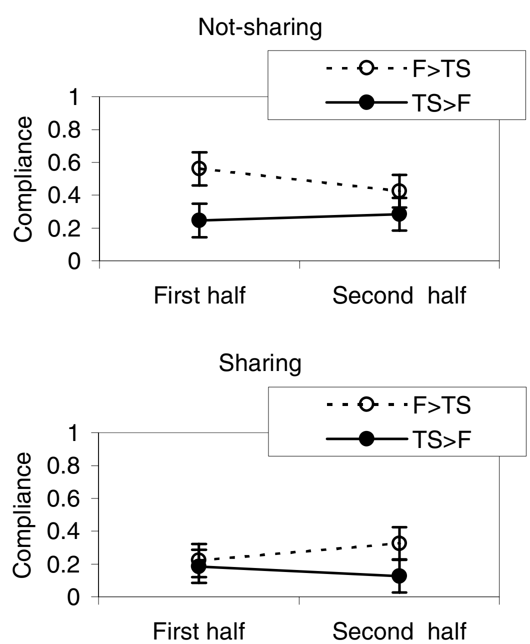

Figure 2. Three-way interaction between sharing information, demand profile, and repetition on compliance.

information decreased participants' compliance with the selected model forecast most in the first half of the experiment when demand was dramatically changing. Interestingly, sharing of information had no effect on the final forecast accuracy. Consistent with reliance, trust ratings also showed a bias towards the FM. The overall mean score of participants' trust was 5.7 for the FM and 5.3 for the TSM. Whether a model failed first (fault order) showed a significant effect on trust in FM, which suggested that people's initial experience with automation might have substantial influence in their subsequent trust (Gao & Lee, in press). Participants' trust in FM was 6.6 when the TSM failed first and only 4.9 when the FM failed first, $F(1, 20) = 12.39$, $p = 0.0022$. Sharing information had no statistically significant effect on trust.

Compared to the overall mean accuracy of initial forecasts, 77%, the overall mean accuracy of final forecasts was 86%. This suggested that using forecasting models improved the forecasting performance. However, there was no significant difference for the final forecast accuracy across any conditions. Note that the mean final forecast accuracy would have been increased to 95% if the participants had relied on the right model and fully complied with the model forecast across all 60 trials.

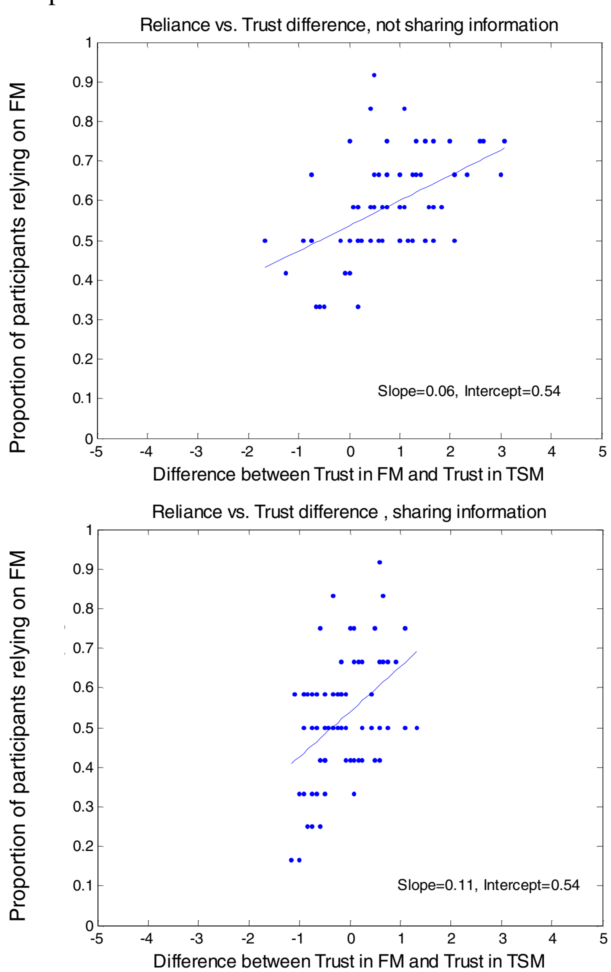

Figure 3. Reliance and trust (Top: not-sharing; Bottom: sharing).

The reliance on FM (or TSM) was plotted versus the difference between model forecast accuracy of FM (or TSM) and TSM (or FM) from the previous trial for not-sharing and sharing groups. A poor correlation between reliance and model reliability was found (r<0.03).

The reliance was plotted versus the difference between trust in FM and TSM for not-sharing and sharing groups, as shown in Figure 3. A linear function was fitted to the data for each. Figure 3 shows a significant difference in the relationship between reliance and trust for the no-sharing and sharing groups. The slope for the sharing group, 0.11, was about twice of the slope for the not-sharing group, 0.06, $p = 0.05$. This suggested that, with shared information, the participants' reliance on the model was more sensitive to the difference in their trust of the two models. With the shared information, there were fewer instances where participants trusted and relied more on FM than on TSM (i.e., the dots on the upper right of Figure 3).

Similarly, the compliance data were plotted as a function of the difference between participants' trust and self-confidence for not-sharing and sharing groups and a linear function was fitted to the data for each. A strong positive linear relationship was found between compliance and the difference of trust and self-confidence, with R-square of 0.87 for the not-sharing and 0.92 for the sharing group. However, there was no significant difference in the slope between the fitting lines for the sharing and no-sharing groups.

DISCUSSION

The participants' reliance depended more on their trust in the model than on the model reliability. The poor correlation between reliance and model reliability confirms that people had considerable difficulty appropriately reacting to the reliability of the forecasting model (Goodwin, 2000; Goodwin & Fildes, 1999; Lawrence & Sim, 1999; Lim & Oconnor, 1996; Lim & O'Connor, 1995). Overall, participants tended to trust more and therefore rely more on the forecasting model in situations where it is more difficult for people to develop accurate judgmental forecasts—those situations in which demand was driven by external factors rather than past history. Participants showed a bias towards FM over TSM. With shared information, this bias was reduced and the reliance depended on trust more strongly (see Figure 3). The strong correlation between compliance and the difference of trust and self-confidence showed again that trust and self-confidence combined to influence people's use of automation (Gao & Lee, in press; Lee & Moray, 1994).

The overall appropriate reliance of 65.6% indicated that participants selected the right model just over half of the time. When a user selected the right model whereas the other user selected the alternative model, the user would see the poor performance of the alternative model via the shared information from the previous trial.

That is, more than half of the time, the shared information exposed relatively poor performance instead of good performance of each model to the participant. The debriefing with participants indicated that the instances of bad performances of models exposed by the shared information made them realize that the model forecasts could be quite wrong even when they also observed some good performances. Therefore they became more conservative, being inclined to depend more on their initial forecasts and less on the model forecasts. It implied that the exposed instances of 'bad performance' of the model drew participants' more attention. These might explain the decreased compliance in the first half of the experiment when demand was dramatically changing (see Figure 2). Regardless of which model was selected, the overall compliance of 0.3 was still far away from the full compliance with the model forecasts of 1.With the decreased compliance, even with the right model selected, the final forecast would not be improved.

In summary, the results showed that people's reliance depended more on their trust than model reliability. People had difficulty in evaluating system reliability and did not comply with the model as they should have when the model was reliable. Model failure during the initial exposure had a strong influence on people's trust in the model. Although sharing information balanced people's bias towards one model over the other, people tended to pay more attention to the exposed instances of 'bad performance' of the model, which might compromise the benefit of sharing information (Dzindolet, Peterson, Pomranky, Pierce, & Beck, 2003). These results suggest that sharing information may promote more appropriate reliance in situations in which people over trust automation, but not in situation in which people tend to under trust automation.

Designers should be cautious about sharing information as well as the way the shared information is presented. This study only considered the situation where two users share the information. Sharing information from multiple sources to provide richer context might reduce users' tendency to focus only on the poor performance instead of overall performance. For example, time is the critical contextual information for demand forecasting. It is therefore important to design the forecasting tool so that the user can be aware of the role of the time frame (e.g., the same change of the demand might be interpreted differently by the user when presented within two different time frames). Presenting the shared information in a comprehensive way might reduce the tendency for a single instance of poor performance to initiate a cycle of diminished trust, subsequently less reliance, leading to further diminished trust. (Gao & Lee, in press). Although not considered as a factor in this study, how the quality and reliability of the shared data influence people's trust and reliance in the forecasting models merits future investigation.

REFERENCES

Berinato, S. (2001, Aug. 1). *What Went Wrong at Cisco*, from http://www.cio.com/archive/080101/cisco.html

Crane, S. (2001). *In the Nick of Time*, from http://www.cfoasia.com/archives/200105-31.htm

Dzindolet, M. T., Peterson, S. A., Pomranky, R. A., Pierce, L. G., & Beck, H. P. (2003). The role of trust in automation reliance. Int. J. Human-Computer Studies 58, 697-718.

Gao, J., & Lee, J. D. (in press). Extending the Decision Field Theory to model operators' reliance on automation in supervisory control situations. *IEEE Transactions on Systems, Man, and Cybernetics*.

Goodwin, P. (2000). Improving the voluntary integration of statistical forecasts and judgment. *International Journal of Forecasting, 16*(1), 85-99.

Goodwin, P., & Fildes, R. (1999). Judgmental forecasts of time series affected by special events: Does providing a statistical forecast improve accuracy? *Journal of Behavioral Decision Making, 12*(1), 37-53.

Jarrett, J. (1990). *Business Forecasting Methods* (Second ed.). 108 Cowley Road, Oxford, OX4 1JF, UK: Basil Blackwell Ltd.

Lawrence, M., Goodwin, P., & Fildes, R. (2002). Influence of user participation on DSS use and decision accuracy. *Omega-International Journal of Management Science, 30*(5), 381-392.

Lawrence, M., O'Connor, M., & Edmundson, B. (2000). A field study of sales forecasting accuracy and processes. *European Journal of Operational Research, 122*(1), 151-160.

Lawrence, M., & Sim, W. (1999). Prototyping a financial DSS. *Omega-International Journal of Management Science, 27*(4), 445-450.

Lee, J. D., & Gao, J. (In press). Trust, automation, and cooperation in supply chains. *Supply Chain Forum: An International Journal, 6*(2), 82-89.

Lee, J. D., & Moray, N. (1994). Trust, self-confidence, and operators' adaptation to automation. *International Journal of Human-Computer Studies, 40*, 153-184.

Lee, J. D., & See, K. A. (2004). Trust in technology: eesigning for appropriate reliance. *Human Factors, 46(1)*, 50-80.

Lim, J. S., & O'Connor, M. (1996). Judgmental forecasting with interactive forecasting support systems. *Decision Support Systems, 16*(4), 339-357.

Lim, J. S., & O'Connor, M. (1995). Judgmental adjustment of initial forecasts - its effectiveness and biases. *Journal of Behavioral Decision Making, 8*(3), 149-168.

Mosier, K. L., Skitka, L. J., Heers, S., & Burdick, M. (1998). Automation bias: Decision making and performance in high-tech cockpits. *International Journal of Aviation Psychology, 8*(1), 47-63.

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors, 39*(2), 230-253.

Paul, K. (March 2002). *Inside Cisco's $ 2 billion Blunder*, from www.business2.com

Rotter, J. B. (1967). A new scale for the measurement of interpersonal trust. *Journal of Personality, 35*(4), 651-665.,

Sterlicchi, J. (2003). *Just Do It? But Not Like Nike - Automating their supply chain management proved to be Nike's achilles heel*, from http://www.ccnmag.com/story.php?id=39